RECEIVED
MAR 22 2002
Technology Center 2600

OIPE JC177
MAR 19 2002
PATENT & TRADEMARK OFFICE

DECLARATION

I, Masanobu Kato, a Japanese citizen of 2312, 7-1, Hikarigaoka 3-chome, Nerima-ku, Tokyo, Japan do hereby declare as follows:

1. I am well acquainted with both the Japanese language and the English language.

2. I personally made the following translation.

3. The following is a true translation into the English language of the specification filed in the Japanese Patent Office on the 20th day of October, 1997 in respect of an application for Patent numbered PCT/JP97/03785.

And I make this solemn declaration conscientiously believing the same to be true.

This 26th day of February, 2002

Masanobu Kato

Masanobu Kato

RECEIVED
MAR 2 2 2002
Technology Center 2600

SPECIFICATION

IMAGE ENCODING DEVICE AND IMAGE DECODING DEVICE

OIPE JC177
MAR 1 9 2002
PATENT & TRADEMARK OFFICE

TECHNICAL FIELD

The present invention relates to an image encoding device and an image decoding device which perform image processing.

BACKGROUND ART

Conventionally, it is always necessary at the decoding side that the analysis of VOP header information be preceded by analyses of a VOP start code, a modulo time base and a VOP time increment which are contained in each VOP header, because no distinction can be made between VOP not to be analyzed (information to be decimated in the case of a low speed shot of an image signal) and VOP to be analyzed (information not to be decimated in the case of a low speed shot of an image signal); hence, there is a problem that the processing involved is cumbersome and low in accuracy.

For decoding and synthesizing encoded signals respectively corresponding to a subject, a background, a log and similar objects which form a pictorial image, it is necessary that each object be added with a synthesizing timing signal (information representing absolute time) necessary for decoding and synthesizing the object. Without the information indicative of absolute time, an image decoding device cannot synthesize the object, and hence it is incapable of image reconstruction. In short, in the case of generating one pictorial image from a plurality of objects including those having no information representative of absolute time, it is impossible with the prior art to combine objects having the required information with those having no such information.

Moreover, the bit length of the modulo time base increases until the next GOV header is multiplexed--this raises a problem that the bit length of the modulo time base keeps on increasing when the GOV header, which is an option, is not multiplexed.

With a view to solving such problems as referred to above, the present invention is to provide an image encoding device and an image decoding device whose processing accuracies improve through simple processing.

Another object of the present invention is to provide an image encoding device and an image decoding device which permit the generation of a pictorial image composed of a plurality of objects based on a time code.

Still another object of the present invention is to prevent the generation of an unnecessary amount of information.

DISCLOSURE OF THE INVENTION

According to the present invention, an image encoding device which encodes an image for each object is provided with: encoding means for encoding the image on the basis of predetermined display speed information; and multiplexing means for multiplexing the encoded image signal, encoded by the encoding means, with the predetermined display speed information prior to transmitting the signal. By this, the display speed information can be sent in a multiplexed form.

Furthermore, according to the present invention, the display speed information is multiplexed for each object. This permits multiplexing the display speed information for each object.

Moreover, according to the present invention, an image decoding device which decodes an encoded bit stream encoded from the pictorial image for each object is provided with: display speed information decoding means for decoding the display speed information from the abovesaid encoded bit stream; and control means for controlling the reconstruction of the image processed for each object on the basis of the display speed information decoded by the display speed information decoding means. This permits smooth and accurate image restoration processing with a simple structure.

Further, according to the present invention, the display speed information is decoded for each object. This provides increased smoothness and increased accuracy in the image

restoration processing with a simple structure.

Still further, according to the present invention, the control means is provided with: decoding time specifying means which specifies the time at which to decode an object, on the basis of the display speed information of the object decoded by the display speed information decoding means and the display speed information of the object preset in the decoding device; and decoding means which decodes the object on the basis of the decoding time specified by the decoding time specifying means. This makes the image restoration processing more smoothly and more accurately with a simple structure.

Still further, according to the present invention, an image encoding device which encodes an image for each object is provided with absolute time multiplexing means by which information representing the absolute time for each object is multiplexed onto said encoded image signal. By this, the information indicating the absolute time can be sent multiplexed form.

According to the present invention, an image decoding device which decodes an encoded bit stream encoded from an image for each object has absolute time analysis means for analyzing the information indicative of the absolute time for each object, and reconstructs the image processed for each object on the basis of the information representing the absolute time analyzed by the absolute time analysis means. This permits simple and accurate image synthesis processing.

According to the present invention, an image encoding device which encodes an image for each object is provided with time information encoding means which encodes, as information defining each image display time for each object, first time information defining the time interval between a reference time and the display time, second time information defining the display time with higher accuracy than by the first time information and the image corresponding to each time; the time information encoding means expresses the first time information as a bit length and, when the bit length of the first time information is longer than a predetermined set value, repeats

a bit shift corresponding to the set value until the bit length becomes shorter than the set value, and at the same time counts the number of repetitions of the bit shift and encodes the number of repetitions of the bit shift and a bit string resulting from the repeated bit shift. This permits reduction of the amount of encoded information to send.

According to the present invention, an image encoding device which encodes an image for each object is provided with time information encoding means which encodes, as information defining each image display time for each object, first time information defining the time interval from a reference time to the display time, second time information defining the display time with higher accuracy than by the first time information and the image corresponding to each time; the time information encoding means has first time information holding means for holding first time information encoded in an image of the immediately previous time, and obtains a bit string corresponding to the difference between first time information of an image to be encoded and first time information of the immediately previously encoded image obtainable from the first time information holding means, and encodes the difference bit string as the first time information of the image to be encoded. This ensures the reduction of the amount of encoded information to be sent.

According to the present invention, an image decoding device which decodes a bit stream encoded from an image for each object is provided with: time information decoding means which decodes, as information defining each image display time for each object, first time information defining the time interval between a reference time and the display time, second time information defining the display time with higher accuracy than by the first time information and the image corresponding to each time; and decoding and synthesizing means which decodes an input encoded image signal for each object and synthesizes the decoded image signals. The time information decoding means decodes, as encoded data of the first time information, the number of repetitions

of a bit shift and a bit string resulting from the repeated bit shift; and the decoding and synthesizing means, which is characterized by the decoding of the first time information by adding the bit string with a code of the length of a predetermined set value by the number of repetitions of the bit shift, synthesizes the decoded image signal on the basis of the first and second time information decoded by the time information decoding means. This permits reception of an image sent with a small amount of encoded information.

According to the present invention, an image decoding device which decodes a bit stream encoded from an image for each object is provided with: time information decoding means which decodes, as information defining each image display time for each object, first time information defining the time interval between a reference time and the display time, second time information defining the display time with higher accuracy than by the first time information and the image corresponding to each time; and decoding and synthesizing means which decodes an input encoded image signal for each object and synthesizes the decoded image signals. The time information decoding means holds first time information of an immediately previously decoded image, and adds a bit string, decoded as first time information of the image to be decoded, with the first time information of the immediately previously decoded image obtainable from the first time information holding means, thereby decoding the first time information of the image to be decoded; and the decoding and synthesizing means synthesizes decoded image signals on the basis of the first and second time information decoded by the time information decoding means. This permits reception of an image sent with a small amount of encoded information.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram depicting the video data structure in MPEG-4; Fig. 2 is a diagram showing a concrete example of VOP; Fig. 3 is a block diagram illustrating a VOP encoder part according to Embodiment 1 of the present invention; Fig. 4 is a block diagram

illustrating an example of the configuration of a header multiplexing part of the VOP encoder part according to Embodiment 1 of the present invention; Fig. 5 is a diagram for explaining a modulo time base and a VOP time increment; Fig. 6 is a block diagram illustrating an example of the configuration of the header multiplexing part of the VOP encoder part according to Embodiment 1 of the present invention; Fig. 7 is a block diagram depicting the internal configuration of a VOP decoder part according to Embodiment 2 of the present invention; Fig. 8 is a block diagram depicting an example of the configuration of a header analysis part of the VOP decoder part according to Embodiment 2 of the present invention; Fig. 9 is a block diagram depicting a system for synthesizing a plurality of objects according to Embodiment 2 of the present invention; Fig. 10 is a block diagram illustrating an example of the configuration of a header analysis part of a VOP decoder part according to Embodiment 3 of the present invention; Fig. 11 is a block diagram illustrating an example of the configuration of the header analysis part of the VOP decoder part according to Embodiment 3 of the present invention; Fig. 12 is a block diagram illustrating an example of the configuration of a header multiplexing part of a VOP encoder part according to Embodiment 4 of the present invention; Fig. 13 is a block diagram illustrating an example of the configuration of a header multiplexing part of a VOP encoder part according to Embodiment 4 of the present invention; Fig. 14 is a block diagram depicting an example of the internal configuration of a VOP decoder part according to Embodiment 5 of the present invention; Fig. 15 is a block diagram depicting an example of the configuration of a header analysis part of the VOP decoder part according to Embodiment 5 of the present invention; Fig. 16 is a block diagram illustrating a system for synthesizing a plurality of objects according to Embodiment 5 of the present invention; Fig. 17 is a block diagram depicting an example of the configuration of the header analysis part of the VOP decoder part according to Embodiment 5 of the present invention; Fig. 18 is a block diagram depicting an example of

the internal configuration of the VOP decoder part according to Embodiment 5 of the present invention; Fig. 19 is block diagram showing an example of the configuration of a header multiplexing part of a VOP encoder part according to Embodiment 6 of the present invention; Fig. 20 is a block diagram illustrating an example of the configuration of a header analysis part of a VOP decoder part according to Embodiment 7 of the present invention; Fig. 21 is a block diagram illustrating an example of the configuration of a header multiplexing part of a VOP encoder part according to Embodiment 8 of the present invention; and Fig. 22 is a block diagram illustrating an example of a header analysis part of a VOP decoder part according to Embodiment 9 of the present invention.

BEST MODE FOR CARRYING OUT THE INVENTION

To facilitate a better understanding of the present invention, a description will be given, with reference to the accompanying drawings, of the best mode for carrying out the invention.

EMBODIMENT 1

In this embodiment, a VOP encoder for the MPEG-4 video encoding system disclosed in ISO/IEC JTC11 SC29/WG11/N1796 will be described which is provided with constituents of this embodiment, i.e. means for encoding an image on the basis of object display speed information and means for multiplexing the display speed information onto an encoded bit stream by adding the information for each object.

The MPEG-4 system is a system that regards a moving picture sequence as a set of moving picture objects taking arbitrary forms temporally and spatially and performs encoding and decoding for each moving picture object. In Fig. 1 there is depicted the video data structure in MPEG-4. In the MPEG-4 the moving picture object containing the time axis is referred to as a video object [Video Object (VO)], a component of the VO as a video object layer {Video Object Layer (VOL)], a component of the VOL as a group of video object plane (Group of Video Object Plane (GOP)], and image data

which represents the state of the GOP at each time and forms the basic unit of encoding as a video object plane [Video Object Plane (VOP)]. The VO corresponds, for example, to each speaker or his background in a video conference scene; the VOL forms the basic unit having inherent temporal and spatial resolutions of the speaker or background; and the VOP is image data of such a VOL at each time (corresponding to a frame). The GOP is a data structure that forms the basic unit for editing a plurality of VOLs or random access thereto; this data structure need not always be used for encoding.

A concrete example of VOP is shown in Fig. 2. In the same figure, there are depicted two VOPs (VOP1 indicating a man and VOP2 a picture on the wall). Each VOP is composed of texture data representing the color gradation level and shape data representing the shape of the VOP. The texture data is composed of a luminance signal of 8 bits per pixel and a color difference signal (of a size subsampled to 1/2 that of the luminance signal in the horizontal and vertical directions), and the shape data is the same binary matrix data as the image size of the luminance signal which sets the inside and outside of the VOP at 0 and 1, respectively.

In the VOP-based moving picture representation a conventional frame image is obtained by arranging a plurality of VOPs in the frame. When the moving picture sequence contains only one VO, each VOP is synonymous with the frame.

In this instance, no shape data exists and only the texture data is encoded.

A description will be given below of an image encoding device of Embodiment 1. This is based on an MPEG-4 video encoder; the MPEG-4 video encoder will hereinafter be referred to as a VOP encoder since it performs encoding for each VOP. Since the operation of the existing VOP encoder is disclosed, for example, in ISO/IEC JTC1/SC29/WG11/N1796, no description will be given of the existing VOP encoder itself, but a description will be given of a VOP encoder that contains constituents of the present embodiment.

Fig. 3 depicts an example of he configuration of the VOP encoder in this embodiment; reference numeral 110 denotes a VOP-to-be-encoded determination part, 111 a shape encoding part, 113 a motion estimation part, 115 a motion compensation part, 118 a texture encoding part, 122 a memory, 124 a header multiplexing part, 126 a video signal multiplexing part, 128 a subtractor, and 129 an adder.

Next, the operation of the encoder will be described. Based on VOP rate information 7 that is set externally or in accordance with the encoding condition, the VOP-to-be-encoded determination part 110 determines the VOP to be encoded in the input object images, and outputs the VOP to be encoded to the shape encoding part 111, the motion estimation part 113 and the subtractor 128. The VOP rate information 7 mentioned herein corresponds to what is called the display speed information in the present invention, and it refers to information that represents how many VOPs in each VOL or GOV are to be displayed per second.

A concrete example of the operation of the VOP-to-be-encoded determination part 110 will be described. When the input object images are 30/sec and the VOP rate information 7 is 15/sec, the VOP-to-be-encoded determination part 110 judges that alternate ones of the VOPs contained in the input object images are to be encoded, and outputs every other VOPs to be encoded.

The VOPs specified by the VOP-to-be-encoded determination part 110 as those to be encoded have their shape data encoded for each area with 16 by 16 pixels, which is commonly called an alpha block, and have their texture data encoded for each area with 16 by 16 pixels which is called a macro block.

The shape encoding part 111 encodes the alpha block input thereinto and outputs encoded shape information 112 and locally decoded shape information 109. The encoded shape information 112 is fed to the video signal multiplexing part 126, whereas the locally decoded shape information 109 is input into the motion estimation part 113, the texture encoding part 115 and the texture encoding part 118. The motion estimation part 113 reads reference data 123a from the memory 122 and performs block

matching for each macro block to obtain motion information 114. In this case, the motion information is obtained by block matching for only the objects contained in the macro block based on the locally decoded shape information 109.

The motion compensation part 115 reads out of the memory 122 reference data 123b on the position indicated by the motion information 114 and generates a predictive image based on the locally decoded shape information 109. The predictive image 116 created in the motion estimation part 115 is input into the subtractor 128 and the adder 129.

The subtractor 128 calculates the difference between the predictive image 116 and the input macro block to provide a prediction-error image 117.

In the texture encoding part 118 the prediction-error image 117 input thereinto is encoded by a predetermined method prescribed by MPEG-4 to obtain encoded texture information 119 and locally-decoded prediction-error image 120. In this instance, only the objects contained in the block are encoded based on the locally decoded shape information 109. The encoded texture information 119 is sent to the video signal multiplexing part 126, and the locally-decoded prediction-error image 120 is output to the adder 129.

The adder 129 adds the predictive image 116 and the locally-decoded prediction-error image 120 to create a decoded image 121, which is written in the memory 122.

In the header multiplexing part 124 respective pieces of header information are multiplexed, and a bit stream 125 obtained by multiplexing the header information is input into the video signal multiplexing part 126.

The video signal multiplexing part 126 multiplexes the encoded shape information 112, the motion information 114 and the encoded texture information 119 onto the bit stream 125 formed by multiplexing respective header information, and outputs an encoded VOP bit stream.

Fig. 4 is a block diagram depicting the configuration of the header multiplexing part shown in Fig. 3. In the same figure,

reference numeral 1 denotes a VO header multiplexing part, 2 a VOL header multiplexing part, 3 a GOV header e multiplexing selection part, 4 a GOV header multiplexing part, 5 a VOP header multiplexing part, 6 GOV multiplexing information, and 7 VOP rate information.

Next, the operation of this embodiment will be described. The VO header multiplexing part 1 creates a bit stream by multiplexing VOP header information, and outputs the bit stream to the VOL header multiplexing part 2.

The VOL header multiplexing part 2 multiplexes VOL header information onto the input bit stream, and outputs the bit stream to the GOV header multiplexing selection part 3.

Based on the GOV multiplexing information 6 indicating whether to perform the multiplexing of the GOV header, the GOV header multiplexing selection part 3 determines the destination of the bit stream fed from the VOL header multiplexing part 2. When the GOV multiplexing information 6 indicates that no multiplexing of the GOV header takes place, the bit stream is output to the VOP header multiplexing part 5, whereas when the GOV multiplexing information 6 indicates that the multiplexing of the GOV header is performed, the bit stream is output to the GOV header multiplexing part 4.

The GOV header multiplexing part 4 outputs the bit stream to the VOP header multiplexing part 5, with the VOP rate information 7 being multiplexed on the inputted bit stream.

Table 1 exemplifies the abovesaid VOP rate information 7, showing four kinds of VOP rates. When the VOP rate is 30/sec, "01" is multiplexed. When the VOP to be encoded is the same as the VOP encoded immediately previously, VOP information "00" is multiplexed but the subsequent VOP header information and VOP data information are not multiplexed. When the VOP rate is variable, VOP rate information "11" is multiplexed.

A VOP start code multiplexing part 8 in the VOP header multiplexing part 5 outputs to a modulo time base (modulo-time-base) multiplexing part 9 and a VOP time increment (VOP-time-increment) multiplexing part 10 a bit stream obtained

by multiplexing a VOP start code onto the input bit stream.

The modulo time base 13 mentioned herein is information that represents what number of seconds will pass until the VOP concerned is displayed after a certain reference time as depicted in Fig. 5, and the VOP time increment 14 is information by which the display time defined by the modulo time base is fine-adjusted with an accuracy of 1/1000th of a second as similarly depicted in Fig. 5. That, is, MPEG-4 permits defining the VOP display time with a precision of 1/1000th of a second.

A management time generating part 12 in the VOP header multiplexing part 5 generates the modulo time base 13 and the VOP time increment 13 based on the VOP rate information 7, and outputs the modulo time base 13 to the modulo time base multiplexing part 9 and the VOP time increment 14 to the VOP time increment multiplexing part 10. When the VOP rate information 7 represents a variable VOP rate, the modulo time base 13 and the VOP time increment 14 are set independently of the VOP rate information 7.

The above-mentioned modulo time base multiplexing part 9 multiplexes the modulo time base 13 onto the bit stream provided from the VOP start code multiplexing part, and outputs the multiplexed bit stream to the VOP time increment multiplexing part 10. The VOP time increment multiplexing part 10 multiplexes the VOP time increment 14 fed thereto from the management time generating part 12 onto the bit stream fed from the modulo time base multiplexing part 9, and outputs the multiplexed bit stream to a video information header multiplexing part 11. The video information header multiplexing part 11 multiplexes a video information header onto the bit stream provided thereto from the VOP time increment multiplexing part 10, and outputs the multiplexed bit stream to the video signal multiplexing part 126.

As described above, according to this embodiment, since the VOP rate information is multiplexed onto the GOP header, a bit stream can be created which enables the decoder side to determine whether or not to require the decoding of the VOP concerned, or to synthesize a plurality of objects, simply by analyzing only

the VOP start code of each VOP header.

It is also possible to define the VOP rate information for each VOL and perform encoding and multiplexing of the VOP rate information as shown in Fig. 6. In this instance, the VOP rate information 7 is determined for each VOL and is multiplexed in the VOL header multiplexing part 2. Based on this, the modulo time base 13 and the VOP time increment 14 are determined.

As described above, the present embodiment has disclosed an example of the image encoding device which encodes images on an objectwise basis and which is provided with encoding means for encoding the images on the basis of predetermined display speed information and multiplexing means for multiplexing the abovesaid predetermined display speed information onto the image signals encoded by the encoding means and for outputting the multiplexed signals.

Furthermore, the present embodiment has disclosed an example of the multiplexing means of the type that multiplexes the above-mentioned display speed information on an object-by-object basis.

EMBODIMENT 2

This embodiment will be described as being applied to a system wherein an image decoding device for decoding from an encoded bit stream the VOP rate information mentioned previously in connection with Embodiment 1 and for outputting it, that is, an MPEG-4 video decoder (hereinafter referred to as a VOP decoder) is provided for each of a plurality of objects and a plurality of decoded objects are synthesized to reconstruct a pictorial image.

A description will be given first of the configuration and operation of the image decoding device (VOP decoder) in this embodiment. Since the operation of the existing VOP decoder is disclosed, for example, in ISO/IEC JTC1/SC29/WG11/N1796, the VOP decoder containing constituents of the present embodiment will be described without referring to the existing VOP decoder itself. The VOP decoder in this embodiment is a decoder that is able to

decode an encoded bit stream generated by the VOP encoder described previously with reference to Embodiment 1.

Fig. 7 depicts an example of the internal configuration of the VOP decoder in this embodiment of the present invention. The VOP decoder is supplied with compressed-encoded data composed of texture data and shape data as described previously with reference to Embodiment 1 and shown in Fig. 2, and decodes the individual pieces of data. In the same figure, reference numeral 150 denotes encoded VOP bit stream, 151 a header analysis part, 152 a bit stream with the header information analyzed, 153 a video signal analysis part, 154 encoded shape data, 155 a shape decoding part, 156 decoded shape data, 157 encoded texture data, 158 motion information, 159 a motion compensation part, 160 predictive texture data, 161 a texture decoding part, 162 decoded texture data, 164 a memory, and 165 reference data.

The operation of the decoder will be described in detail with reference to the same figure. The encoded VOP bit stream 150 is input into the header analysis part 151, wherein the header information is analyzed following a predetermined syntax. The bit stream having the header information analyzed in the header analysis part 151 is fed into the video signal analysis part 153, wherein it is analyzed into the encoded shape data 154, the encoded texture data 157 and the motion information 158. The shape decoding part 155 decodes the encoded shape data input thereinto, and outputs the decoded shape data 156.

The motion compensation part 159 generates the predictive texture data 160 from the reference data 165 in the memory 164 and the motion information 158 provided from the video signal analysis part 153. Based on the encoded texture data 157 and the predictive texture data 160, the texture decoding part 161 reconstructs image data by the method prescribed in MPEG-4, generating the decoded texture data 162. The decoded texture data 162 is used for subsequent VOP decoding, and hence it is written in the memory 164.

Fig. 8 depicts the internal configuration of the header analysis part 151 that is characteristic of this embodiment of

the present invention. In the same figure, reference numeral 51 denotes a start code analysis part, 52 a VO header analysis part, 53 a VOL header analysis part, 54 GOV header analysis part, 58 VOP rate information, and 55 a VOP header analysis part. The header analysis part 151 in this embodiment is characterized in that the GOV header analysis part 54 decodes the VOP rate information of VOP contained in the GOV concerned from the bit stream and outputs to the outside. A description will be given later of how to use the VOP rate information 58.

The start code analysis part 51 analyzes the start code contained in the encoded VOP bit stream 150 input thereinto. A bit stream is output to the VO header analysis part when the analyzed start code is indicative of VOL, to the VOL header analysis part 53 when the analyzed start code is indicative of VOL, to the GOV header analysis part 54 when the analyzed start code is indicative of GOV, and to the VOP header analysis part 55 when the analyzed start code is indicative of VOP. Incidentally, upon completion of the analysis in the VOP header analysis part 55, the bit stream is output to the video signal analysis part 153.

The VO header analysis part 52 analyzes VO header information from the input bit stream, and outputs the resulting bit stream to the start code analysis part 51. The VOL header analysis part 53 analyzes VOL header information from the input bit stream, and outputs the resulting bit stream to the start code analysis part 51. The GOV header analysis part 54 analyzes GOV header information from the input bit stream, and outputs the resulting bit stream to the start code analysis part 51. At this time, the VOP rate information 58 contained in the GOV header information is decoded and output. The VOP header analysis part 55 analyzes VOP header information from the input bit stream, and outputs the resulting bit stream via the start code analysis part 51 to the video signal analysis part 153.

With the VOP decoder of the above configuration of operation, it is possible to output, for each GOV, the VOP rate information of VOPs contained therein. Fig. 13 illustrates a system that uses

this information to synthesize a plurality of objects.

In the same figure, reference numeral 200 denotes an encoded VOP bit stream a, 201 an encoded VOP bit stream b, 202 an encoded VOP bit stream c, 203a a VOP decoder part for decoding the encoded VOP bit stream 200 a200, 203b a VOP decoder part for decoding the encoded VOP bit stream b201, 203c a VOP decoder part for decoding the encoded bit stream c202, 204 a decoded object image a, 205 a decoded object image b, 206 a decoded object image c, 207 VOP rate information a, 208 VOP rate information b, 209 VOP rate information c, 210 a composition part, and 211 a decoded pictorial image. The decoded object image herein mentioned refers to an image that is obtained by combining the decoded shape data 154 and the corresponding decoded texture data 162 for each of VOPs and then integrating such combined pieces of data for each group of VOPs (for example, GOV or VOL).

The encoded VOP bit streams a200 to c202 are decoded by the VOP decoder parts 203a to 203c corresponding thereto, respectively, by which the decoded VOP images a204 to c206 are generated. At this time, the VOP decoder parts decode the corresponding VOP rate information a207 to c209, and output them to the composition part 210. Based on the VOP rate information a207 to c209, the composition part 210 determines the time of the frame where to synthesize the decoded VOP images in the decoded image 211, and maps them into the frame corresponding to the determined time. Let it be assumed, for example, that the decoded image 211 is displayed at a rate of 30 video object planes per sec (which corresponds to a ordinary TV signal display speed). Furthermore, assume the following situations.

The decoded VOP image a204 is displayed at a rate of 5/sec (that is, the VOP rate information a207 indicates the 5/sec rate).

The decoded VOP image b205 is displayed at a rate of 10/sec (that is, the VOP rate information indicates the 10/sec rate).

The decoded VOP image c206 is displayed at a rate of 15/sec (that is, the VOP rate information c209 indicates the 15/sec rate).

In this instance, the decoded VOP images a204 to c206 are

all mapped into the first image frame at each second in the decoded image 211; the decoded VOP image a204 is mapped into every five image frames including the first at each second; the decoded VOP image b205 is mapped into every 10 image frames including the first at each second; and the decoded VOP image c206 is mapped into every 15 images frames including the first at each second. By this, it is possible to display a pictorial image with a plurality of objects synthesized in the image frames in accordance with their display speeds.

As described above, by using VOP decoder which decodes the encoded bit stream having the VOP rate information encoded in the GOV layer, a system which synthesizes a plurality of object into a reconstructed image can be implemented with a simple structure.

The VOP rate information may also be encoded for each VOL at the image encoding device side. In this case, it is possible, at the image decoding device side, to decode the VOP rate encoded for each VOL and easily synthesize a plurality of objects for each VOL as described above.

While this embodiment employs the VOP decoder as a system for synthesizing a plurality of objects, it is also feasible to use only one VOP decoder for a system that decodes only one object to reconstruct an image.

As described above, according to this embodiment, the image decoding device which decodes the bit stream encoded from an image on an object-by-object basis is provided with display speed information decoding means for decoding display speed information from the encoded bit stream and control means for controlling the reconstruction of the image encoded on the object-by-object basis through utilization of the display speed information decoded by the display speed information decoding means.

Furthermore, in this embodiment the display speed information decoding means has been described to decode the display speed information object by object.

EMBODIMENT 3

This embodiment is directed to another modification of the VOP decoder described above in Embodiment 2. The VOP decoder according to this embodiment has a function of specifying the VOP to be decoded on the basis of the value of the VOP rate that the decoder assumes.

Since the VOP decoder of this embodiment differs from Embodiment 2 only in the configuration and operation of the header analysis part 151, a description will be given only in this respect.

Fig. 10 is a block diagram illustrating the configuration of the header analysis part of the VOP decoder part according to Embodiment 3, in which the VOP rate at the encoder side and the VOP rate at the decoder side do not match. In the figure, reference numeral 59 denotes a VOP-to-be-decoded selection part, which compares a VOP rate from the GOV header analysis part 54 and a VOP rate assumed at the decoder side, and outputs VOP select information 62. And the VOP header analysis part 55 has a counter 60 has a counter part 60 in addition to a time management information header analysis part 56 and a video information header analysis part 57.

Next, the operation of this embodiment will be described. The VOP-to-be-decoded selection part 59 outputs to the counter part 60 of the VOP header analysis part 55 the VOP select information that indicates information about the VOP to be decoded according to the result of comparison between the VOP rate 58 analyzed in the GOV header analysis part 54 and the VOP rate 61 assumed at the decoder side. The counter part 60 uses the VOP select information 62 to determined whether to decode the VOP header information that follows the VOP start code contained in the input bit stream.

More specifically, when the VOP rate 58 analyzed in the GOV header analysis part 55 is 30 planes/sec and the VOP rate assumed at the decoder side is 15 planes/sec, the VOP select information 62 indicating that every other VOPs are analyzed is output to the counter part 60 in the VOP header analysis part 55. The

counter part 60 first counts every VOP header input thereinto by a counter 60a.

Then, based on the count value input thereinto from the counter 60a and the VOP rate select information 62 from the VOP-to-be-decoded selection part 59, decision means 60b decides whether the input VOP needs to be analyzed. When it is decided that the input VOP needs to be analyzed, the input bit stream is output to the time management information header analysis part 56. When it is decided that the input VOP need not be analyzed, the input bit stream is output to the start code analysis part 51.

A concrete example will be described below. When the VOP rate select information 62 is one that one VOP needs to be analyzed for every three VOPs, the decision means 60b judges that the VOP needs to be analyzed for which the count value from the counter 60a can be divided by 3 without a remainder, and that the VOP need not be analyzed for which the count value from the counter 60a is divided by 3, with a remainder of 1 or 2.

Incidentally, while the VOP decoder of this embodiment has been described to be adapted for use in the case where the VOP rate information is contained in the GOV header, the VOP rate information may also be contained in the VOL header as described previously with reference to Embodiment 2. In such an instance, the VOL header analysis part 300 needs only to be equipped with the function of decoding the VOP rate information 58 as shown in Fig. 11.

Moreover, the VOP decoder of this embodiment can be used not only in a system which synthesizes a plurality of objects but also in a system which decodes and reconstructs only one object.

As described above, the decoder according to this embodiment has control means which is provided with: decoding time specifying means for specifying the time when to decode an object on the basis of the object display information decoded by the display speed information decoding means and the object display speed information preset in the decoding device; and decoding means for decoding the object at the decoding time specified by

the decoding time specifying means.

EMBODIMENT 4

This embodiment is directed to another example of the VOP encoder described previously in Embodiment 1. The VOP encoder of this embodiment has a function of adding, for each VOL, the time code that defines the absolute display time of each VOP contained in the VOL concerned.

Here, the time code mentioned herein is time information disclosed in IEC standard publication 461 for "time and control codes for video tape recorders", which is information that defines the display time of an image at each time forming a moving picture (a frame in MPEG-2 and a VOP in MPEG-4) with an accuracy of hour/minute/second. For example, in the case of performing video editing on a frame-by-frame basis by commercial video editor, the addition of this information to each frame makes it possible to access a desired frame simply by designating the value of the time code.

Since the VOP encoder of this embodiment differs from the encoder of Embodiment 1 only in the configuration and operation of the header multiplexing part 124, a description will be given in this respect alone.

Fig. 12 is a block diagram illustrating the configuration of the header multiplexing part of the VOP encoder part according to Embodiment 4 of the present invention; the parts identical with those in Embodiments 1 of Fig. 4 are marked with the same reference numerals as in the latter, and no description will be repeated.

Next, the operation of this embodiment will be described. The bit stream with the VO header information multiplexed thereon in the VO header multiplexing part 1 is input into the VOL header multiplexing part 2. The VOL header multiplexing part 2 multiplexes on the input bit stream the VOL header information and a time code 18 forming the basis for time management, and outputs the bit stream to the GOV header multiplexing selection part 3.

The GOV header multiplexing selection part 3 determines the destination of the input bit stream from the VOL header multiplexing part 2 on the basis of the GOV multiplexing information 6 indicating whether to perform the multiplexing of the GOV header. When the GOV multiplexing information 6 indicates that no multiplexing of the GOV header is performed, the bit stream is output to the VOP header multiplexing part 5, whereas when the GOV multiplexing information 6 indicates that the multiplexing of the GOV header is performed, the bit stream is output to the GOV header multiplexing part 4. In this instance, the GOV header multiplexing part 4 multiplexes the GOV header information on the bit stream fed from the GOV header multiplexing selection part 3, and outputs the bit stream to the VOP header multiplexing part 5.

The VOP header multiplexing part 5 multiplexes the VOP start code, the time management information header and the video information header onto the input bit stream, and outputs it to the video signal multiplexing part 126 (see Fig. 3). Incidentally, the operations of the video signal multiplexing part 126 and the parts following it are the same as described above.

As described above, according to this embodiment, since the time code is multiplexed onto the VOL header which is always encoded in MPEG-4, it is possible to form a bit stream which permits the creation of a pictorial image composed of a plurality of objects on the basis of the time code. Moreover, in the case of performing edits while decoding the encoded bit stream according to this embodiment by a commercial object-by-object video editor, a VOP at an arbitrary time of objects can freely be accessed randomly at all times. These effects provide increased flexibility in image synthesis.

Incidentally, while the encoder of this embodiment has been described to add the time code for each VOL, the encoder may also be configured to add the time code information for each VOP. This could be implemented by such a configuration as shown in Fig. 13 in which the time code 18 defining the absolute display time

of each VOP is input into and multiplexed by a VOP header multiplexing part 301.

Furthermore, this embodiment has been described to involve the encoding of the VOP rate information, but it is a matter of course that the multiplexing of the time is independent of the VOP rate information, and even when the VOP rate information is not encoded, the same effects as mentioned above are obtainable.

As described above, the image encoding device of this embodiment which encodes images on the object-by-object basis is provided with absolute time multiplexing means by which information representing the absolute time of each object is multiplexed onto an encoded image signal.

EMBODIMENT 5

This embodiment will be described in connection with a system which a plurality of VOP decoders each of which decodes and outputs a time code from the VOL header in the encoded bit stream and synthesizes a plurality of decoded objects into an image.

A description will be given first of the configuration and operation of the VOP decoder in this embodiment. The internal configuration of the VOP decoder in this embodiment is depicted in Fig. 14. Since this decoder differs from the VOP decoder of Embodiment 2 only in the configuration and operation of a header analysis part 302, a description will be given below in this respect alone. The header analysis part 302 has a function of decoding and outputting the time code in the VOL header.

Fig. 15 illustrates the internal configuration of the header analysis part 302. In the same figure, reference numeral 303 denotes a VOL header analysis part. The start code analysis part 51 analyzes the start code contained in the input encoded VOP bit stream 150. The start code analysis part outputs the bit stream to the VO header analysis part 52 when the analyzed start code indicates VO, to the VOL header analysis part 303 when the analyzed start code indicates VOL, to the GOV header analysis part 54 when the analyzed start code indicates GOV, and to the VOP header analysis part 55 when the analyzed start code indicates

VOP. Incidentally, upon completion of the analysis in the VOP header analysis part 55, the bit stream is output therefrom to the video signal analysis part 153.

The VO header analysis part 52 analyzes the Vo header contained in the input bit stream, and outputs the analyzed bit stream to the start code analysis part 51. The VOL header analysis part 303 analyzes the VOL header information in the input bit stream, and outputs the analyzed bit stream to the start code analysis part 51. In this case, the time code 64 contained in the VOL header information is decoded and output. The GOV header analysis part 54 analyzes the GOV header information in the input bit stream, and outputs the analyzed bit stream to the start code analysis part 51. The VOP header analysis part 55 analyzes the VOP header information in the input bit stream, and outputs the analyzed bit stream via the start code analysis 51 to the video signal analysis part 153.

With the VOP decoder of the above configuration and operation, it is possible to output, for each VOL, the absolute display time of each VOP contained therein. In Fig. 16 there is depicted a system which uses this information to synthesize a plurality of objects.

In the same figure, reference numeral 400 denotes an encoded VOP bit stream a, 401 an encoded VOP bit stream b, 402 an encoded bit stream c, 403 a VOP decoder part for decoding the encoded VOP bit stream a400, 403b a VOP decoder part for decoding the encoded VOP bit stream b401, 403c a VOP decoder part for decoding the encoded VOP bit stream c402, 404 a decoded object image c, 405 a decoded object image b, 406 a decoded object image c, 407 a time code a, 408 a time code b, 409 a time code c, 410 a composition part, and 411 a decoded image. What is intended to mean by the decoded object image is an image obtained by combining the decoded shape data 156 and the corresponding decoded texture data 162 for each of VOPs and then integrating such combined pieces of data for each group of VOPs (for example, GOV or VOL).

The encoded VOP bit stream a400 to the encoded VOP bit stream c402 are decoded by the VOP decoder parts 403a to 403c

corresponding thereto, respectively, by which the decoded VOP images a404 to c406 are generated. At this time, the VOP decoder parts decode the corresponding time codes a407 to c409, and output them to the composition part 210. Based on the time codes a407 to c409, the composition part 210 determines the time of the frame of the decoded image 411 where to synthesize the decoded VOP of each decoded object image, and maps them into the frame corresponding to the determined time. For example, assume the following situations.

* The composition part has a time code generation capability, and determines the absolute display time of each image frame to synthesize.

* Assume that 01:00:00 is decoded as the time code of the first VOP of the decoded object image a404, where 01:00:00 represents (hour):(minute):(second).

* Assume that 01:00:10 is decoded as the time code of the first VOP of the decoded object image b405.

* Assume that 01:01:00 is decoded as the time code of the first VOP of the decoded object image c406.

Assuming that the time code of the first image frame of the decoded image 411 defined in the composition part 410 is 01:00:00, the decoded object image a404 is mapped into the first frame of the decoded image 411, the decoded object image b405 is mapped 10 seconds after the first frame of the decoded image 411, and the decoded object image c406 is mapped one minute after the first frame of the decoded image 411; thus, an object of displaying the decoded objects can be performed. By this, it is possible to display a pictorial image with a plurality of video objects synthesized in the image frames in correspondence to the reference absolute times.

As described above, by using the VOP decoder which decodes the encoded bit stream having the time code encoded in the GOV layer, a system which synthesizes a plurality of object into a reconstructed image can be implemented with a simple structure.

The time code may also be encoded for each VOL at the image encoding device side as depicted in Fig. 17. In this case, it

is possible, at the image decoding device side, to decode the time code encoded for each VOL and synthesize a plurality of objects for each VOL as described above.

It is also possible to configure such a VOP decoder as shown in Fig. 18 in which an encoded bit stream with the VOP rate multiplexed thereon are input into the VOL header, together with the time code. In this instance, since the absolute display time of the first VOP of the VOL is determined by the time code and then the absolute display time of each VOP can easily be detected from the VOP rate information, it is feasible to configure a system that synthesizes a plurality of objects with more ease.

While this embodiment employs the VOP decoder as a system for synthesizing a plurality of objects, it is also possible to use only one VOP decoder in a system that decodes only one object to reconstruct an image.

As described above, according to this embodiment, the image decoding device which decodes the bit stream encoded from an image on an object-by-object basis is provided with absolute time analysis means for analyzing, for each object, information indicating the absolute time therefor and means for reconstructing the image processed on the object-by-object basis through utilization of the information indicating the absolute time analyzed by the absolute time analysis means.

EMBODIMENT 6

This embodiment will be described in connection with an improved modulo time base encoding method and a VOP encoder therefor which are used to represent the modulo time base (corresponding to first time information) and the VOP time increment (corresponding to second time information) which are now used in MPEG-4.

A description will be given first of the method for representing the modulo time base 20 in MPEG-4.

As described previously in Embodiment 1, the value of the modulo time base is information that indicates what number of seconds will pass until the VOP concerned is displayed after a

certain reference time as shown in Fig. 5, and the information expresses the number of seconds in terms of the number of bits of the value "1." The end of the data is clearly indicated by adding the value "0." That is, when the display is provided after 5 seconds, the information becomes "111110." With this method, when the reference time does not change at all, the amount of information of the modulo time base increases infinitely. At present, MPEG-4 defines the reference time by the time code that is multiplexed onto the GOV header, but since the GOV header is a option, the GOV header need not always be encoded under MPEG-4 prescriptions. That is, there is a fear that the value of the modulo time base becomes longer limitlessly unless the GOV header appears. This embodiment implements an encoder that obviates such a problem in encoding the data of the modulo time base.

This embodiment requires modifying only the configuration and operation of the header multiplexing part 124 of the VOP encoders described so far, a description will be focused on this part alone.

Fig. 19 illustrates the internal configuration of the header multiplexing part 124 in this embodiment of the present invention. Reference numeral 500 denotes a VOP header multiplexing part , 19 a bit length calculating part, 20 a modulo time base, 21 a shifted modulo time base, 22 an information bit indicating a repeat count, and 501 a modulo time base.

Next, the operation of this embodiment will be described. The bit stream with the VO header information multiplexed thereon in the VO header multiplexing part 1 is input into the VOL header multiplexing part 2. The VOL header multiplexing part 2 multiplexes the VOL header information onto the input bit stream, and outputs the multiplexed bit stream to the GOV header multiplexing selection part 3.

The GOV header multiplexing selection part 3 determines the destination of the bit stream from the VOL header multiplexing part 2 on the basis of the GOV multiplexing information 6 indicating whether to perform multiplexing of the GOV header. When the GOV multiplexing information 6 indicates that no

multiplexing of the GOV header is performed, the bit stream is output to the VOP header multiplexing part 5, whereas when the GOV multiplexing information 6 indicates that the multiplexing of the GOV header is performed, the bit stream is output to the GOV multiplexing part 4. In this case, the GOV header multiplexing part 4 multiplexes the GOV header information onto the bit stream from the GOV header multiplexing selection part 3, and outputs the multiplexed bit stream to the VOP header multiplexing part 5.

The VOP start code multiplexing part 8 in the VOP header multiplexing part 500 multiplexes the VOP start code onto the input bit stream, and outputs the multiplexed bit stream to the modulo time base multiplexing part 501. The bit length calculating part 19 in the VOP header multiplexing part 500 compares the bit length of the modulo time base 20 and a present positive threshold value; when the bit length of the modulo time base 20 is longer than the threshold value, the modulo time base 20 is left-shifted repeatedly by the length of the threshold value until the bit length of the modulo time base becomes shorter than the threshold value, and the modulo time base 21, which is the resulting bit string, and the information bit 22, which indicates the shift-repeat count, are output. The information bit 22 indicating the shift-repeat count may be provided as a binary number that expresses the shift-repeat count by a predetermined number of bits, or as a variable bit length that expresses the shift-repeat count by a variable-length code.

A concrete example of the operation in the bit length calculation part 19 will be described below. With the abovesaid threshold value set at 4, if the modulo time base 20 is "1111111110," the shift-repeat count is two and the shifted modulo time base 21 is "10." If expressed by a fixed-length two bits, the information bit 22 indicating the shift-repeat count is "10."

The modulo time base multiplexing part 501 in the VOP header multiplexing part 500 multiplexes onto the bit stream from the VOP start code multiplexing part 8 the shifted modulo time base

21 and the information bit 22 indicating the shift-repeat count, and outputs the multiplexed bit stream to the VOP time increment multiplexing part 10.

The VOP time increment multiplexing part 10 multiplexes the VOP time increment onto the bit stream from the modulo time base multiplexing part 10, and outputs the multiplexed bit stream to the video information header multiplexing part 11. The video information header multiplexing part 11 multiplexes the video information header onto the bit stream from the VOP time increment multiplexing part 10, and outputs the multiplexed bit stream to the video signal multiplexing part 26.

As described above, according to this embodiment, the modulo time base is expressed by two kinds of information bits (the shifted modulo time base and the information bit indicating the shift-repeat count), and these two kinds of information bits are multiplexed instead of multiplexing the modulo time base expressed as prescribed in MPEG-4 at present; hence, it is possible to suppress the amount of information generated as compared with the method according to MPEG-4.

As described above, the image encoding device of this embodiment which encodes images on the object-by-object basis is provided with time information encoding means which encodes, as information defining the display time of an image at each time on the object-by-object basis, first time information defining the time interval between the reference time and the display time and second information defining the display time with a higher accuracy than that of the time defined by the first time information and the image corresponding to each time; the time information encoding means expresses the first time information by conversion into a bit length, and when the bit length of the first time information is longer than a predetermined set value, a bit shift corresponding to the set value is repeated until the bit length becomes shorter than the set value, and at the same time, the number of bit shifts is counted, then the shift-repeat count and the bit string obtained by the repetitions of the bit shift are encoded.

EMBODIMENT 7

The present embodiment is directed to a VOP decoder which decodes the modulo time base information multiplexed onto the encoded bit stream in the modulo time base multiplexing part described above in Embodiment 6 and uses the decoded information and the VOP time increment to define the display time of each VOP.

Since Embodiment 10 differs from the VOP decoders described so far only in the configuration and operation of the header analysis part 151, a description will be given in this respect alone.

Fig. 29 illustrates the internal configuration of the header analysis part 151 in this embodiment of the present invention. Reference numeral 502 denotes a VOP header analysis part, 65 a modulo time base analysis part, 66 a VOP time increment analysis part, 67 a modulo time base calculation part, 69 a shifted modulo time base, and 70 an information bit indicating a shift-repeat count.

Next, the operation of this embodiment will be described. The start code analysis part 51 analyzes the start code contained in an encoded bit stream having multiplexed thereon the input shifted modulo time base and the information bit indicating the shift-repeat count, and outputs the bit stream 152 to the VO header analysis part 52 when the analyzed start code is contained in the VO header, to the VOL header analysis part 53 when the analyzed start code is contained in the VOL header, to the GOV header analysis part 54 when the analyzed start code is contained in the GOV header, to the VOP header analysis part 55 when the analyzed start code is contained in the VOP header, and to the video signal analysis part 153 (see Fig. 7) when the analyzed start code is contained in the VOP data information. The operations of the video signal analysis part and the parts following it are the same as described so far.

The modulo time base analysis part 65 in the VOP header analysis part 502 analyzes the shifted modulo time base 69 and

the information bit 70 indicating the shift-repeat count contained in the bit stream fed from the start code analysis part 51, and outputs the shifted modulo time base 69 and the information bit 70 indicating the shift-repeat count to the modulo time base calculation part 67 and the bit stream to the VOP time increment analysis part 66.

The modulo time base calculation part 67 calculates the modulo time base from the shifted modulo time base 69 and the information bit 70 indicating the shift-repeat count, and outputs it to the composition part 210. More specifically, the value of the modulo time base is restored by reversing the procedure described previously with reference to Embodiment 9. In the case where a preset positive threshold value (The decoder side also required to set exactly the same value as the threshold value described in respect of the encoder of Embodiment 9.) and the shifted modulo time base 69 is "10" and the information bit 70 indicating the shift-repeat count is "10," "1111111110" with "11111111" added to the high-order bit of "10" is the restored value of the modulo time base. The thus obtained restored value of the modulo time base is used to define the display time of the VOP concerned, together with the VOP time increment information.

The VOP time increment analysis part 66 analyzes the VOP time increment contained in the bit stream fed from the modulo time base analysis part 65, and outputs the analyzed bit stream to the video information header analysis part 57. The video information header analysis part 57 analyzes the video information header contained in the bit stream fed from the VOP time increment analysis part 66, and outputs the analyzed bit stream to the video signal analysis part 153.

As described above, the decoder of this embodiment is configured to calculate the modulo time base from the two kinds of information bits (the shifted modulo time base and the information indicating the shift-repeat count); hence it is possible to analyze the bit stream described later in Embodiment 9 which has a smaller amount of information generated than that

by the encoded representation prescribed in MPEG-4.

As described above, the image display device of this embodiment which decodes a bit stream with images encoded on the object-by-object basis is provided with: time information decoding means which decodes, as information defining the display time of an image at each time on the object-by-object basis, first time information defining the time interval between the reference time and the display time and second information defining the display time with a higher accuracy than that of the time defined by the first time information and the image corresponding to each time; and decoding and synthesizing means for decoding the input encoded image signal on the object-by-object basis and synthesizing these decoded image signals. The time information decoding means the time information encoding means expresses the first time information by conversion into a bit length, and when the bit length of the decodes, as encoded data of the first time information, a bit string derived from the bit-shift repeat count and the repeated bit shift and decodes the first time information by adding the bit string with code of a length of the predetermined set value by the bit-shift repeat count, and the decoding and synthesizing means synthesizes the decoded image signal on the basis of the first and second time information decoded by the time information decoding means.

EMBODIMENT 8

The present embodiment will be described in connection with another improved modulo time base encoding method and a VOP encoder therefor which are used to represent the modulo time base and the VOP time increment which are now used in MPEG-4.

Since this embodiment differs from the VOP encoders described so far only in the configuration and operation of the header multiplexing part 124, a description will be given in this respect alone.

Fig. 21 illustrates the internal configuration of the header multiplexing part 124 in Embodiment 11. Reference numeral 503 denotes a VOP header multiplexing part, 23 a modulo time base

holding part, 24 a difference modulo time base generating part, 25 a difference modulo time base multiplexing part, and 26 a difference modulo time base.

The VOP start code multiplexing part 8 in the VOP header multiplexing part 503 multiplexes the VOP start code onto the input bit stream, and outputs the multiplexed bit stream to the difference modulo time base multiplexing part 25.

The modulo time base holding means 23 in the VOP header multiplexing part 503 holds the value of the modulo time base of the immediately previously encoded VOP, and after modulo time base of the immediately preceding encoded VOP is output therefrom, the modulo time base of the VOP to be encoded is written in the modulo time base holding part.

The difference modulo time base generating part 24 in the VOP header multiplexing part 503 calculates a bit string of the difference between the modulo time base of the immediately preceding encoded VOP input thereinto from the modulo time base holding part 23 and the modulo time base of the VOP to be decoded, then calculates the difference modulo time base 26 based on the number of bits "1" contained in the calculated difference bit string, and outputs it to the difference modulo time base multiplexing part 25.

Now, a concrete example of the generation of the difference modulo time base will be described.

In the case where the modulo time base of the immediately previously encoded VOP is "11110" (decimal numeral: 30) and the modulo time base of the VOP to be encoded is "111110" (decimal numeral: 62), the difference bit string becomes "100000" (decimal numeral: 32). Then, the number of bits "1" contained in the thus calculated difference bit string "100000" is one. In the case of calculating the difference modulo time base by such a conversion table as Table 2, the difference modulo time base corresponding to one bit "1" is "10," and consequently, "10" is output as the difference modulo time base. Table 2 is an example of the conversion table, and other conversion tables may also be defined.

Also it is possible to obtain the difference modulo time base simply by making a comparison of bit lengths alone. For example, in the above example the bit length of the modulo time base of the immediately previously encoded VOP is 5 and the bit length of the modulo time base of the VOP to be encoded is 6; therefore, a value of 1 is obtained as the difference. By using this value as a substitute for the "number of bits "1" contained in the difference bit string" in Table 2, the difference modulo time base can be expressed.

The difference modulo time base multiplexing part 25 in the VOP header multiplexing part 503 multiplexes the difference modulo time base 26 onto the input bit stream, and outputs the multiplexed bit stream to the VOP time increment multiplexing part 10.

The VOP time increment multiplexing part 10 in the VOP header multiplexing part 503 multiplexes the VOP time increment onto the bit stream fed from the difference modulo time base multiplexing part 25, and outputs the multiplexed bit stream to the video information header multiplexing part 11.

As described above, the encoder according to this embodiment is adapted to express the modulo time base as the difference modulo time base and multiplex the difference modulo time base instead of encoding the modulo time base in the form presently prescribed in MPEG-4; hence, the amount of information generated can be made smaller than in the case of using the method prescribed in MPEG-4.

As described above, the image encoding device of this embodiment which encodes images on the object-by-object basis is provided with time information encoding means which encodes, as information defining the display time of an image at each time on the object-by-object basis, first time information defining the time interval between the reference time and the display time and second information defining the display time with a higher accuracy than by the first time information and the image corresponding to each time; the time information encoding means has first time information holding means for holding the first time information encoded for the image at the immediately

preceding time, and calculates a bit string of the difference between the first time information of the image to be encoded and the first time information of the image at the immediate preceding time provided from the first time information holding means, and encodes the difference bit string as the first time information of the image to be encoded.

EMBODIMENT 9

The present embodiment is directed to a VOP decoder which restores the value of the modulo time base of the VOP concerned from information about the difference modulo time base multiplexed onto the encoded bit stream in the difference modulo time base multiplexing part 25 described above in Embodiment 11 and uses the restored modulo time base value to define the display time of each VOP.

Since this embodiment differs from the VOPs described so far only in the configuration and operation of the header analysis part 151, a description will be given in this respect alone.

Fig. 22 illustrates the internal configuration of the header analysis part 151 in this embodiment of the present invention. Reference numeral 504 denotes a VOP header analysis part, 71 a difference modulo time base analysis part, 72 a modulo time base generating part, 73 a VOP time increment analysis part, 74 a modulo time base holding part, and 75 a difference modulo time base.

The difference modulo time base analysis part 71 in the VOP header analysis part 504 analyzes the difference modulo time base 75 contained in a bit stream fed from the start code analysis part 51, and outputs the analyzed difference modulo time base 75 to the modulo time base generating part 72 and the analyzed bit stream to the VOP time increment analysis part 73.

The modulo time base generating part 72 in the VOP header analysis part 504 calculates the number of bits "1" contained in the bit string of the difference between the modulo time base of the immediately previously analyzed VOP and the modulo time base of the VOP to be analyzed, from the analyzed difference modulo time base 75 on the basis of the conversion table depicted as

Table 3, then generates a modulo time base from the calculated number of bits "1" and the modulo time base of the immediately previously analyzed VOP available from the modulo time base holding part 74, and outputs the thus generated modulo time base to the modulo time base holding part 74.

A concrete example of the generation of the modulo time base will be described. Assume that the analyzed difference modulo time base is "10" and that the modulo time base analyzed immediately previously and held in the modulo time base holding part is "11110." In the case of calculating from the conversion table shown in Table 3 the number of bits "1" contained in the bit string of the difference between the modulo time base of the immediately previously analyzed VOP and the modulo time base of the VOP to be analyzed, it is known that the number of bits "1" contained in the difference bit stream corresponding to the difference modulo time base "10" is one. Then, one bit "1" is added to the most significant bit of the modulo time base "11110" of the immediately previously analyzed VOP to obtain a modulo time base. The conversion table of Table 2 is an example, and other conversion tables may also be defined and used. The restored value of the modulo time base is used to define the display time of the VOP concerned, together with the VOP time increment information.

Furthermore, the "number of bits "1" contained in the bit string of the difference between the modulo time base of the immediately previously analyzed VOP and the modulo time base of the VOP to be analyzed" may also be a bit stream encoded as the "difference value between the bit length of the modulo time base of the immediately previously analyzed VOP and the bit length of the modulo time base of the VOP to be analyzed," and in this case the interpretation of such a conversion table as Table 2 needs only to be changed.

The modulo time base holding part 74 in the VOP header analysis part 504 holds the modulo time base of the immediately previously analyzed VOP, and after modulo time base of the immediately preceding encoded VOP is output therefrom, the modulo

time base of the VOP to be encoded is input into the modulo time base holding part.

The VOP time increment analysis part 73 in the VOP header analysis part 504 analyzed the VOP time increment contained in the bit stream fed from the difference modulo time base analysis part 71, and outputs the analyzed bit stream to the video information header analysis part 57.

As described above, the decoder of this embodiment is adapted to calculate the modulo time base from the difference time modulo base with a small amount of information; hence it is possible to analyze the bit stream described previously in Embodiment 8 which has a smaller amount of information generated than that by the encoded representation prescribed in MPEG-4.

As described above, the image decoding device of this embodiment which decodes a bit stream with images encoded on the object-by-object basis is provided with: time information decoding means which decodes first time information defining the time interval between the reference time and the display time and second information defining the display time with a higher accuracy than that of the time defined by the first time information, as information defining the display time of an image at each time in an image series, and the image corresponding to each time; and decoding and synthesizing means for decoding the input encoded image signal on the object-by-object basis and synthesizing these decoded image signals. The time information decoding means holds the first time information of the immediately previously decoded image, then adds the first time information of the immediately decoded image available from the first time information holding means to a bit string decoded as the first time information of the image to be decoded, thereby decoding the first time information of the image to be decoded; and the decoding and synthesizing means synthesizes the decoded image signal on the basis of the first and second time information decoded by the time information decoding means.

EMBODIMENT 10

While in this embodiment there have been described above that the image encoding device multiplexes the display speed information onto the encoded image signal and that the image encoding device multiplexes the absolute time information onto the encoded image signal, it is also possible to implement an image encoding device which multiplexes both the display speed information and the absolute time information onto the encoded image signal.

This can be done by a parallel or series arrangement of display speed information multiplexing means and absolute time information multiplexing means in the respective image encoding device described in each of the above-described embodiments.

The same goes for the image decoding device side. To put it simply, there have been described above in Embodiments 1 through 12 that the image decoding device decodes the display speed information and uses this decoded display speed information to reconstruct images processed on the object-by-object basis and that the image decoding device decodes the absolute time information and uses the decoded absolute time information to reconstruct images processed on the object-by-object basis; however, it is also possible to implement an image decoding device which reconstructs the images processed for each object on the basis of the display speed information and the absolute time information.

This can be done by a parallel or series arrangement of the display speed information decoding part and the absolute time information decoding part in the respective image decoding device described in each of the above-mentioned embodiments so that images processed for each object are reconstructed based on the information decoded in each decoding part.

With the above configuration, the image restoration and synthesis can be performed more smoothly and more accurately.

EMBODIMENT 11

While in this embodiment there have been described above that the image encoding device encodes and multiplexes the display

speed information on the encoded image signal and that the image encoding device multiplexes the first time information, the second time information and the image, it is also possible to implement an image encoding device which encodes and multiplexes the display speed information, the first time information, the second time information and the image.

This can be done by a parallel or series arrangement of the display speed information multiplexing means and the first and second time information and image multiplexing means in the respective image encoding device described in each of the above-mentioned embodiments.

The same goes for the image decoding device side. To put it briefly, there have been described above in this embodiment that the image decoding device decodes the display speed information and, based on the decoded display speed information, reconstructs images processed for each object and that the image decoding device decodes the first time information, the second time information and the image and, based on the decoded first time information, second time information and image, reconstructs the image; however, it is also possible to implement an image decoding device reconstructs images on the basis of the display speed information and the decoded first and second time information.

This can be done by a parallel or series arrangement of the display speed information decoding part and the time information decoding part in the respective image decoding device described in each of the above-described embodiments so that images processed for each object are reconstructed based on the information decoded in each decoding part (means).

With the above configuration, the image restoration can be performed more smoothly and more accurately with a small amount of coded information sent.

EMBODIMENT 12

While in the above-described embodiments there have been described above that the image encoding device multiplexes the

absolute time information and encoded image signal and that the image encoding device encodes and multiplexes the first time information, the second time information and the image, it is also possible implement an image encoding device which encodes and multiplexes the absolute time information, the first and second time information and the image.

In addition, this also can be achieved by a parallel or series arrangement of the absolute time multiplexing means and the first and second time information and image encoding and multiplexing means in the respective image encoding device described in each of these embodiments.

On the other hand, the same goes for the image decoding device side. To put it simply, there have been described above in the present embodiments that the image decoding devices decode the absolute time information and, based on the decoded absolute time information, reconstruct images processed for each object and that the image decoding devices decode the first time information, the second time information and the image and reconstruct the image, based on the decoded first time information, second time information and image; however, it is also possible to implement an image decoding device reconstructs images on the basis of the absolute time information and the decoded first and second time information.

Further, this can also be achieved by a parallel or series arrangement of the absolute time information decoding part and the time information decoding part in the respective image decoding device described in each of the above-mentioned embodiments so that images processed for each object are reconstructed based on the information decoded in each decoding part (means).

With the above configuration, the image restoration can be achieved more smoothly and more accurately with a small amount of coded information sent.

INDUSTRIAL APPLICABILITY

As described above, according to the present invention, the

image decoding device analyzes the display speed information multiplexed in the image encoding device and performs decoding based on the analyzed display speed information, thereby permitting smooth image reconstruction with a simple structure. Furthermore, the image decoding device decodes the absolute time information multiplexed in the image encoding device and performs decoding based on the analyzed absolute time information, thereby permitting the image reconstruction with ease and with high accuracy. Moreover, the image decoding device decodes the first and second time information encoded in the image encoding device and decodes the input image signal based on the decoded first and second time information, thereby permitting the reception of the image signal with a small amount of information sent.

Table 1

VOP Rate	VOP Rate Information
30/sec	01
15/sec	10
Still Picture	00
Variable	11

Table 2

Number of Bits "1" Contained in Difference Bit String	Difference Modulo Time Base
0	0
1	10
2	110
...	...
n	11 ... 10

"1" continues for n bits